**USED CARS PRICE PREDICTION**

**(Data Wizards)**

**TEAM MEMBERS:**

CHIRUDEEP GORLE -016682627

RAPARLA SRAVANI-016656601

KANCHETI SAI PRAGNA- 016698552

**GITHUB LINK :**

[**https://github.com/Sravani-05/257-ML-Project.git**](https://github.com/Sravani-05/257-ML-Project.git)

**1.ABSTRACT**

The manufacturer sets the price of a new automobile in the industry, with the government incurring some additional expenditures in the form of taxes. Customers who purchase a new automobile may rest confident that the money they spend will be well spent. However, due to rising new automobile prices and buyers' inability to afford them, used car sales are on the rise all over the world. As a result, there is a pressing need for a Used Car Price Prediction system that accurately evaluates the car's worthiness based on several factors. The current system involves a procedure in which a vendor sets a price at random, and the buyer has no knowledge of the automobile or its current worth.In truth, the vendor has no clue what the car's current value is or what price he should sell it for. To address this issue, we have devised a methodology that will be extremely effective. Regression algorithms are employed because they produce a continuous value rather than a classified value as an output. As a result, rather than predicting the price range of a car, it will be possible to forecast the exact price of a car. A user interface has also been created that collects input from any user and shows the price of a car based on their inputs.

**2.INTRODUCTION**

Due to the numerous elements that influence the price of a used vehicle on the market, determining if the advertised price of a used automobile is accurate is a difficult process. The goal of this project is to create machine learning models that can reliably estimate the price of a used automobile based on its characteristics so that clients can make an educated decision. On a dataset consisting of the selling prices of various brands and models, we implement and evaluate various learning algorithms. We will compare the performance of various machine learning algorithms like Linear Regression, Ridge Regression, Lasso Regression, Elastic Net, Decision Tree Regressor and choose the best out of it. The car's pricing will be determined based on a number of factors. Regression Algorithms are used because they give us a continuous number as an output rather than a categorized value, allowing us to forecast the real price of a car rather than its price range. A user interface has also been designed that collects input from any user and shows the price of a car based on their inputs.

**3.Literature Review**

We will examine several applications and approaches that motivated us to create our project in this chapter. We conducted a background survey to determine the core concepts of our project and then utilized those concepts to collect information such as the technological stack, algorithms, and flaws in our project, enabling us to develop a better product.

**Gaadi.com**

CarDekhoGaadi entered the Indian used vehicle market to enhance the process of selling a car for you. From determining the best resale price for a used automobile to conducting loan settlement and RC transfer, They've got you covered. This way, you can sit back and relax as we utilize a combination of cutting-edge technology and seasoned personnel to find you the greatest deal on your vehicle.

**4.Problem Identification & objectives**

* To create a supervised machine learning model for predicting a vehicle's worth based on numerous factors.
* The system being developed must be feature-based, which means that feature-wise prediction must be achievable.
* Providing graphical comparisons for a clearer picture.

To create an efficient and effective model that estimates the price of a used automobile based on the inputs of the user. To attain high precision. To create a user-friendly User Interface (UI) that takes user input and forecasts the pricing.

**5.System Methodology**

In the system, there are 2 major phases:

1. Training phase: Using the data in the data set, the system is trained to fit a model (line/curve) depending on the algorithm used.

2. Testing phase: the system is given inputs and its functionality is tested.

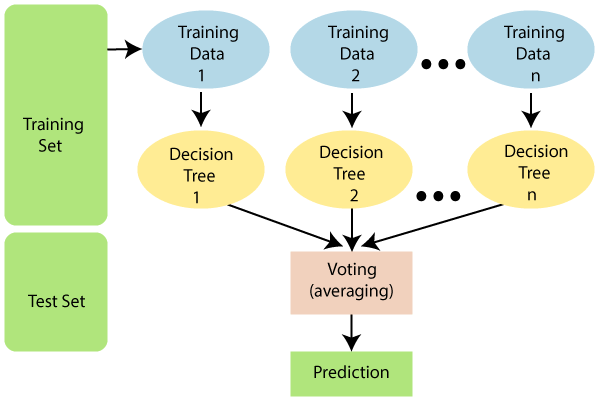
The accuracy is verified. As a result, the data that is used to train or test the model must be appropriate. Because the system is meant to identify and estimate the price of used automobiles, separate algorithms must be utilized to accomplish the two goals.

Different algorithms were assessed for their accuracy before being chosen for future usage.The well-suited one for the task was chosen.

**6.Overview of technologies**

**Random forest :**

A random forest is a data construct applied to machine learning that develops large numbers of random decision trees analyzing sets of variables. This type of algorithm helps to enhance the ways that technologies analyze complex data.



* It takes less training time as compared to other algorithms.
* It predicts output with high accuracy, even for the large dataset it runs efficiently.
* It can also maintain accuracy when a large proportion of data is missing.

**WORKING:**

**Step-1:** Select random K data points from the training set.

**Step-2:** Build the decision trees associated with the selected data points (Subsets).

**Step-3:** Choose the number N for decision trees that you want to build.

**Step-4:** Repeat Step 1 & 2.

**HTML:**

HTML is the standard markup language for Web pages.

With HTML you can create your own Website.

HTML is easy to learn

**CSS:**

CSS stands for Cascading Style Sheets

CSS describes how HTML elements are to be displayed on screen, paper, or in other media

CSS saves a lot of work. It can control the layout of multiple web pages all at once

External stylesheets are stored in CSS files

**JAVASCRIPT:**

JavaScript is used to develop interactive web applications. JavaScript can power featured like interactive images, carousels, and forms. The language can be used with back-end frameworks like Node. js to power the mechanics behind a web page, such as form processing and payments.

It is a lightweight, interpreted, or just-in-time compiled programming language with first-class functions.

JavaScript is a fun and flexible programming language. It's one of the core technologies of web development and can be used on both the front-end and the back-end.

**Python:**

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL).

Python is a high-level, interpreted, interactive and object-oriented scripting language.

**Characteristics of Python:**

Following are important characteristics of Python Programming −

It supports functional and structured programming methods as well as OOP.

It can be used as a scripting language or can be compiled to byte-code for building large applications.

It provides very high-level dynamic data types and supports dynamic type checking.

It supports automatic garbage collection.

It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

**Flask:**

Flask is a popular Python web framework, meaning it is a third-party Python library used for developing web applications.

It is used for developing web applications using python, implemented on Werkzeug and Jinja2.

Advantages of using Flask framework are: There is a built-in development server and a fast debugger provided.

**7.Implementation**

**Web Page:**

**index.html:**

<!DOCTYPEhtml>

<htmllang="en">

<head>

    <metacharset="UTF-8">

    <metaname="viewport"content="width=device-width, initial-scale=1.0">

    <linkrel="stylesheet"href="style.css">

    <title>Document</title>

</head>

<body>

    <divstyle="color:blue">

        <formaction="{{ url\_for('predict')}}"method="post">

            <h2>Predictive analysis</h2>

            <h3>Year</h3>

            <inputid="first"name="Year"type="number ">

            <h3>What is the Showroom Price?(In lakhs)</h3><br>

            <inputid="second"name="Present\_Price"required="required">

            <h3>How Many KilometersDrived?</h3>

            <inputid="third"name="Kms\_Driven"required="required">

            <h3>How much owners previously had the car(0 or 1 or 3) ?</h3><br>

            <inputid="fourth"name="Owner"required="required">

            <h3>What Is the Fuel type?</h3><br>

            <selectname="Fuel\_Type\_Petrol"id="fuel"required="required">

                <optionvalue="Petrol">Petrol</option>

                <optionvalue="Diesel">Diesel</option>

                <optionvalue="Diesel">CNG</option>

            </select>

            <h3>Are you A Dealer or Individual</h3><br>

            <selectname="Seller\_Type\_Individual"id="resea"required="required">

                <optionvalue="Dealer">Dealer</option>

                <optionvalue="Individual">Individual</option>

            </select>

            <h3>Transmission type</h3><br>

            <selectname="Transmission\_Mannual"id="research"required="required">

                <optionvalue="Mannual">Manual Car</option>

                <optionvalue="Automatic">Automatic Car</option>

            </select>

            <br><br>

            <buttonid="sub"type="submit ">Calculate the Selling Price</button>

            <br>

        </form>

        <br><br><h3>{{ prediction\_text }}<h3>

    </div>

</body>

</html>

**Style.css:**

body {

            background-image: url('jakob-rosen-vnwJ9r2ab34-unsplash.jpg');

            background-size: 100%100%;

            text-align: center;

            padding: 0px;

        }

        #research {

            font-size: 18px;

            width: 100px;

            height: 23px;

            top: 23px;

        }

        #box {

            border-radius: 60px;

            border-color: 45px;

            border-style: solid;

            font-family: cursive;

            text-align: center;

            background-color: rgb(168, 131, 61);

            font-size: medium;

            position: absolute;

            width: 700px;

            bottom: 9%;

            height: 850px;

            right: 30%;

            padding: 0px;

            margin: 0px;

            font-size: 14px;

        }

        #fuel {

            width: 83px;

            height: 43px;

            text-align: center;

            border-radius: 14px;

            font-size: 20px;

        }

        #fuel:hover {

            background-color: rgb(241, 200, 184);

        }

        #research {

            width: 99px;

            height: 43px;

            text-align: center;

            border-radius: 14px;

            font-size: 18px;

        }

        #research:hover {

            background-color: rgb(241, 200, 184);

        }

        #resea {

            width: 99px;

            height: 43px;

            text-align: center;

            border-radius: 14px;

            font-size: 18px;

        }

        #resea:hover {

            background-color: rgb(241, 200, 184);

        }

        #sub {

            width: 120px;

            height: 43px;

            text-align: center;

            border-radius: 14px;

            font-size: 18px;

        }

        #sub:hover {

            background-color: rgb(241, 200, 184);

        }

        #first {

            border-radius: 14px;

            height: 25px;

            font-size: 20px;

            text-align: center;

        }

        #second {

            border-radius: 14px;

            height: 25px;

            font-size: 20px;

            text-align: center;

        }

        #third {

            border-radius: 14px;

            height: 25px;

            font-size: 20px;

            text-align: center;

        }

        #fourth {

            border-radius: 14px;

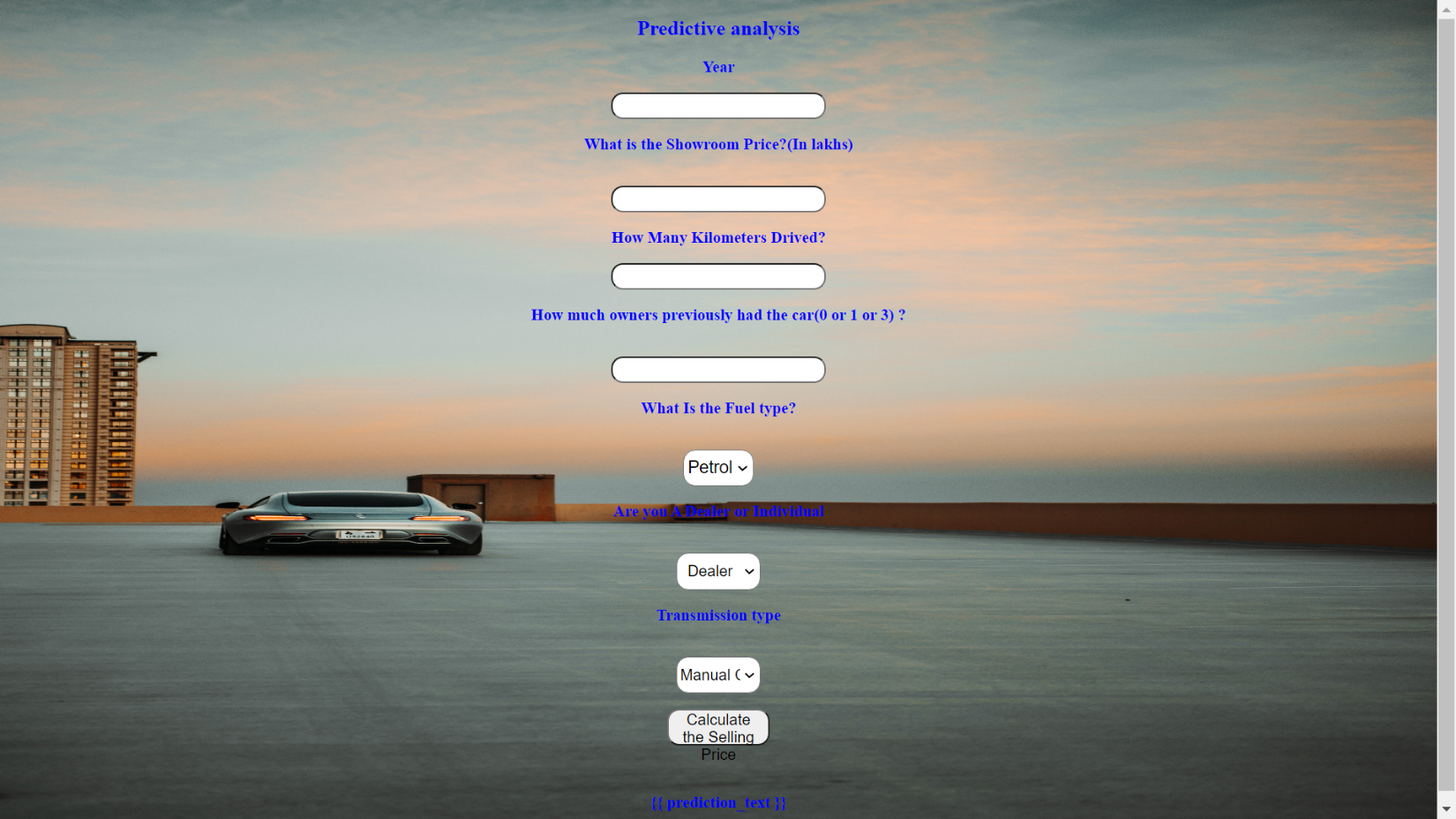
            height: 25px;

            font-size: 20px;

            text-align: center;

        }

**Output:**

****

**Backend-Flask:**

**App.py:**

****

****

**ML Model:**

**1.Coding:**

* Setting a virtual environment
* When creating an end-to-end project, this should be the first step. We require a new virtual environment since each project necessitates a distinct set and version of libraries. By creating a separate environment for each project, we can feed all of the necessary libraries to that environment. To do so, take these steps...

conda create -n carprediction python=3.6  
#some essential package for the environment will be installed #automatically then you will get option  
[y/n] ---> y #click y

>> pip install jupyter notebook # installing jupyter notebook on env  
>>jupyter notebook

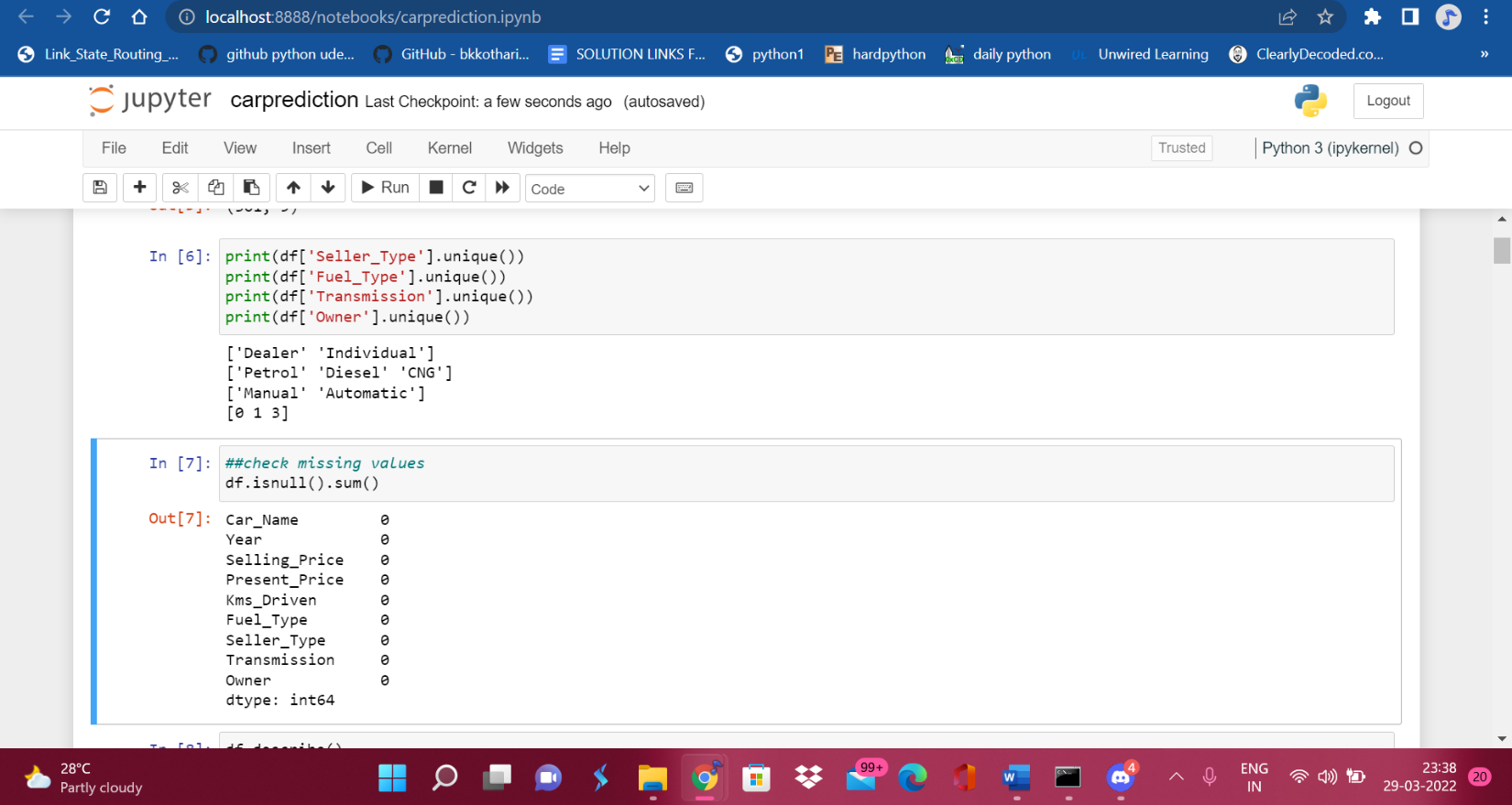
**Acquiring data set and importing libraries:**

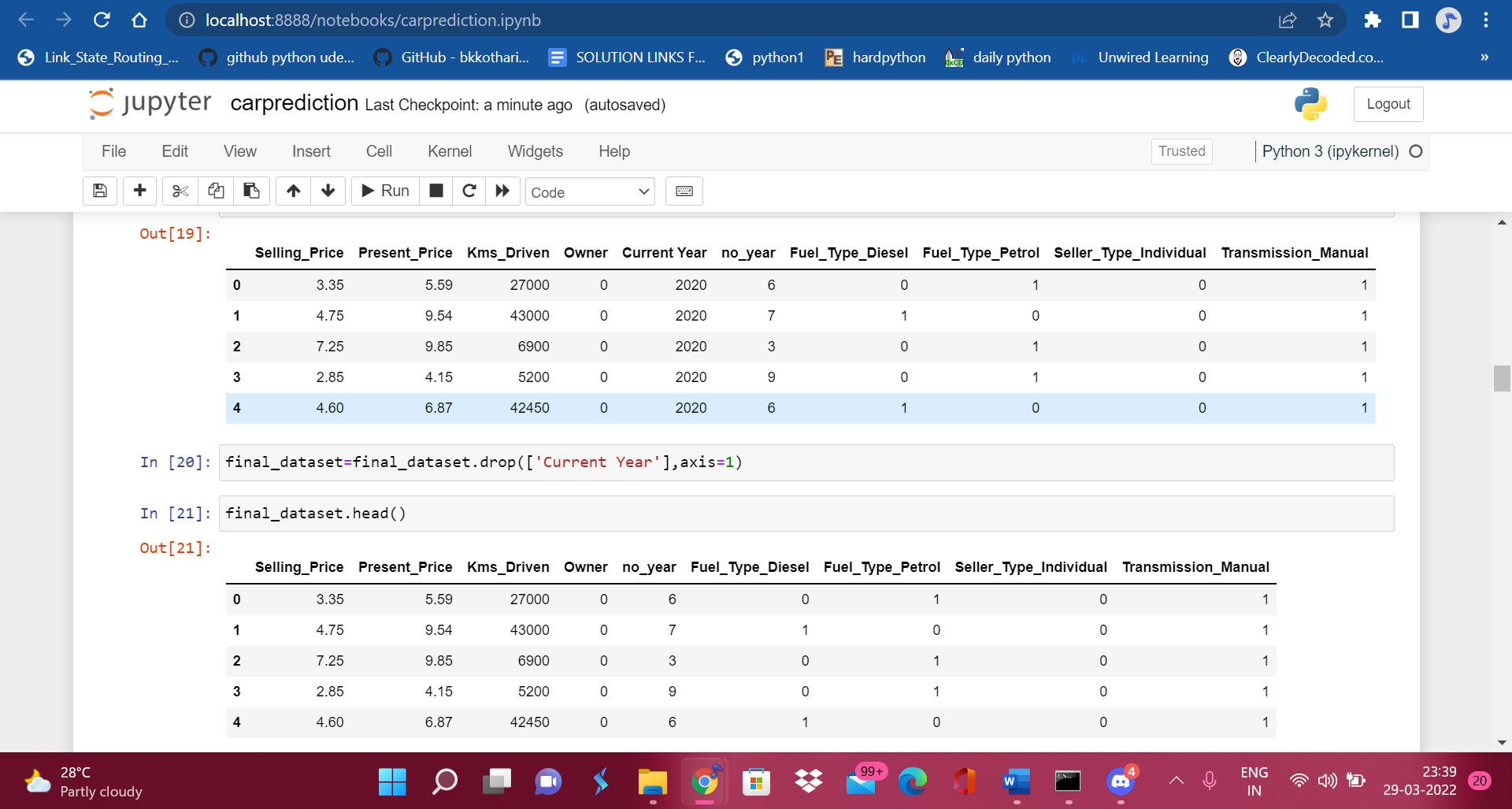
We have downloaded the data set from here, which is in csv format. Now we are going to import all of the necessary libraries for this project.

import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
from sklearn.preprocessing import StandardScaler  
from sklearn.model\_selection import train\_test\_split

**Data analysis:**

Data analysis is a process of inspecting , cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.



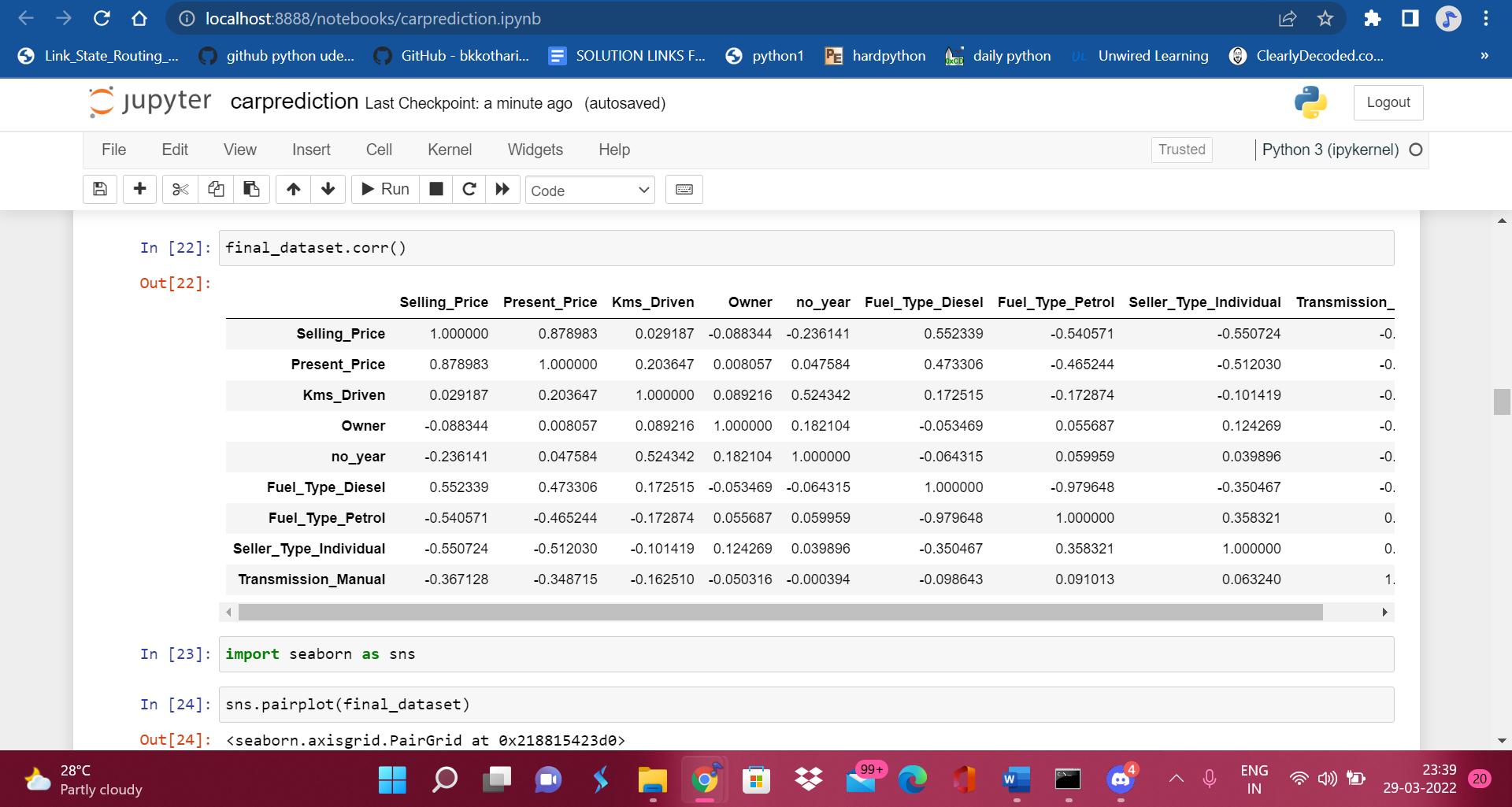


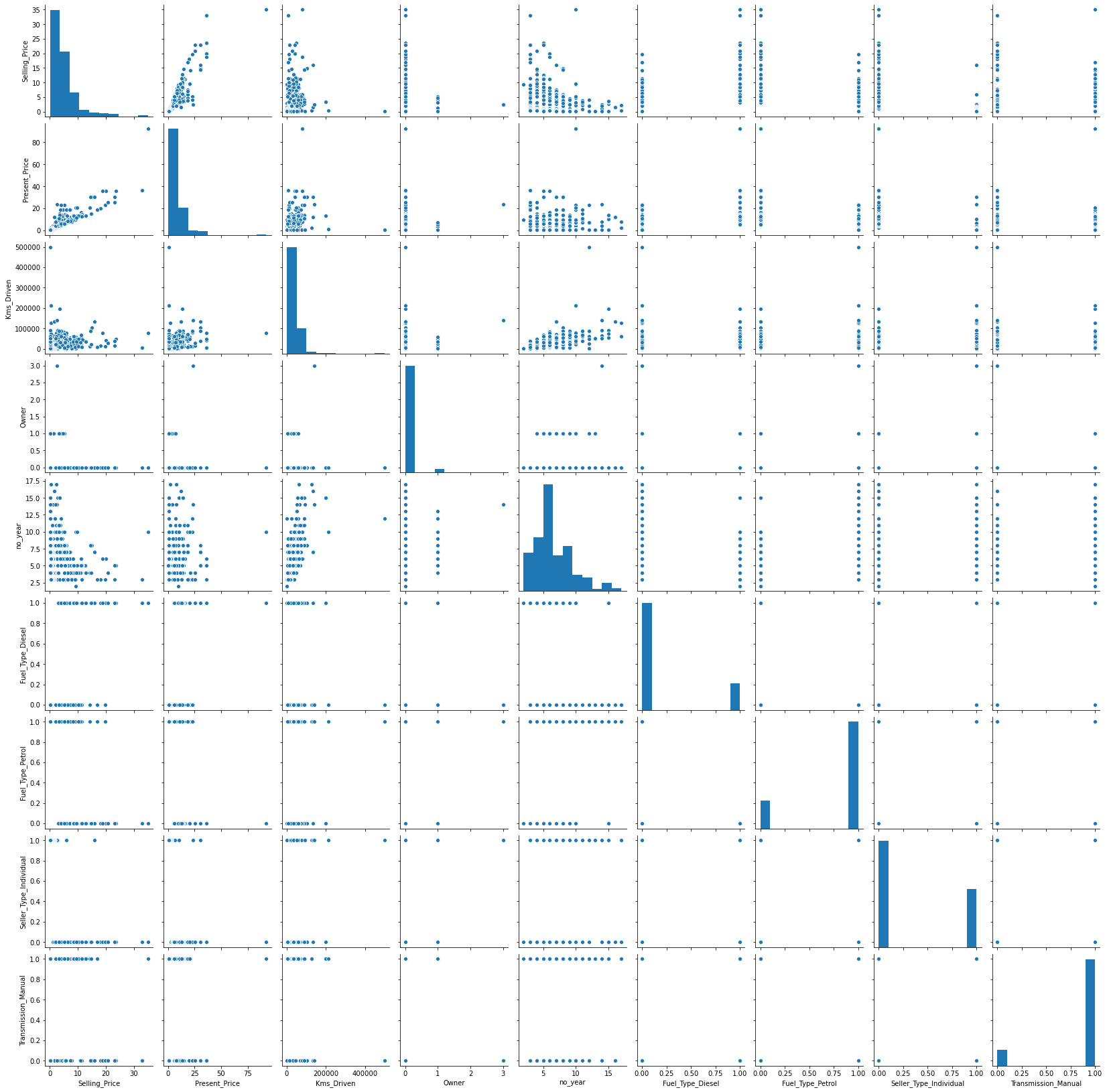
**Data visualization:**

**plotting pair plot:**

Because we can't see multidimensional scatter plots, we need to use pair plots to see each and every dimension of (Dimension with numerical variable)multidimensional data.

A pairplot plots a pairwise relationship in a dataset. The pairplot function creates a grid of axes such that each variable in data will be shared in the y-axis across a single row and in the x-axis across a single column.

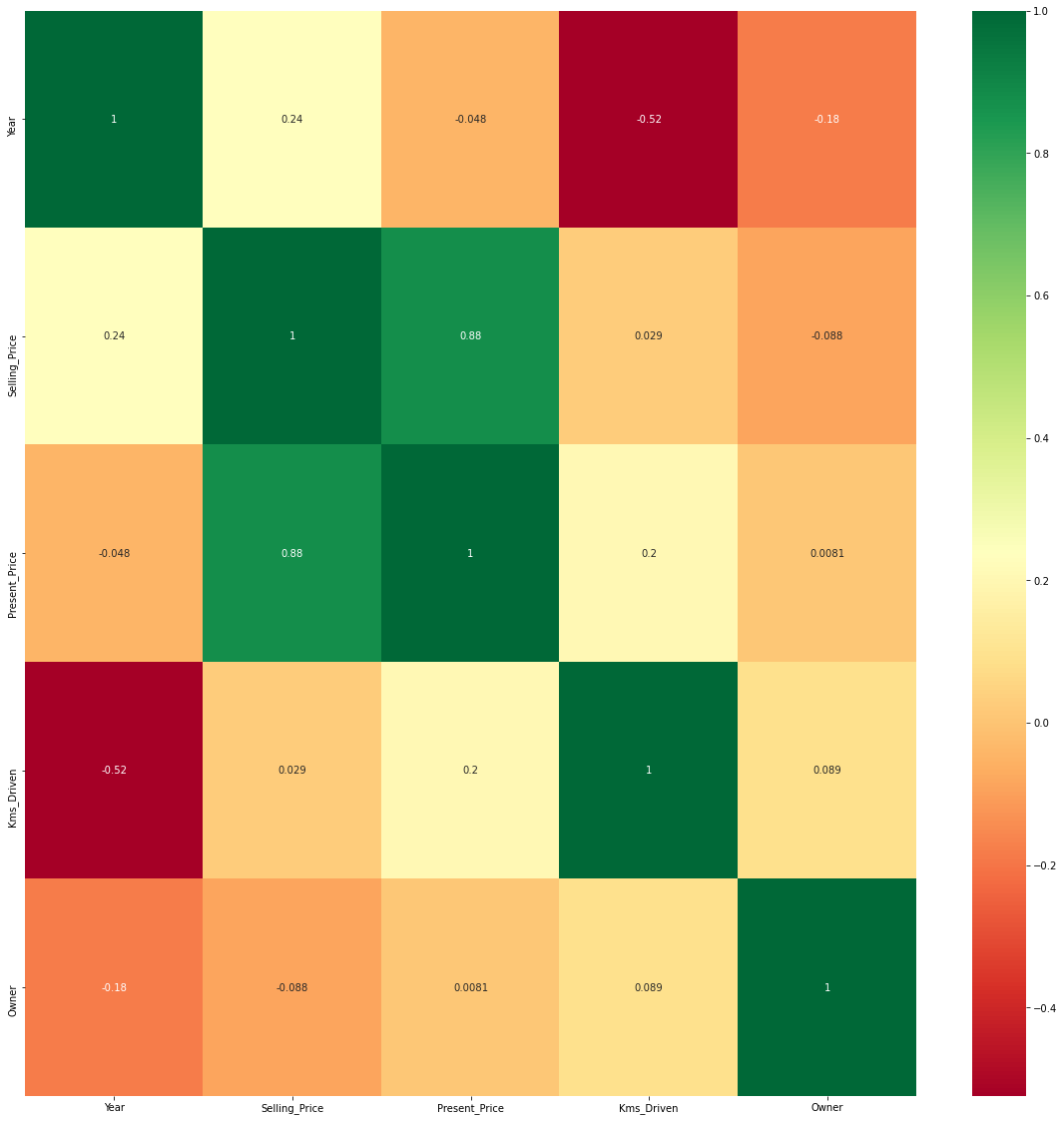
****

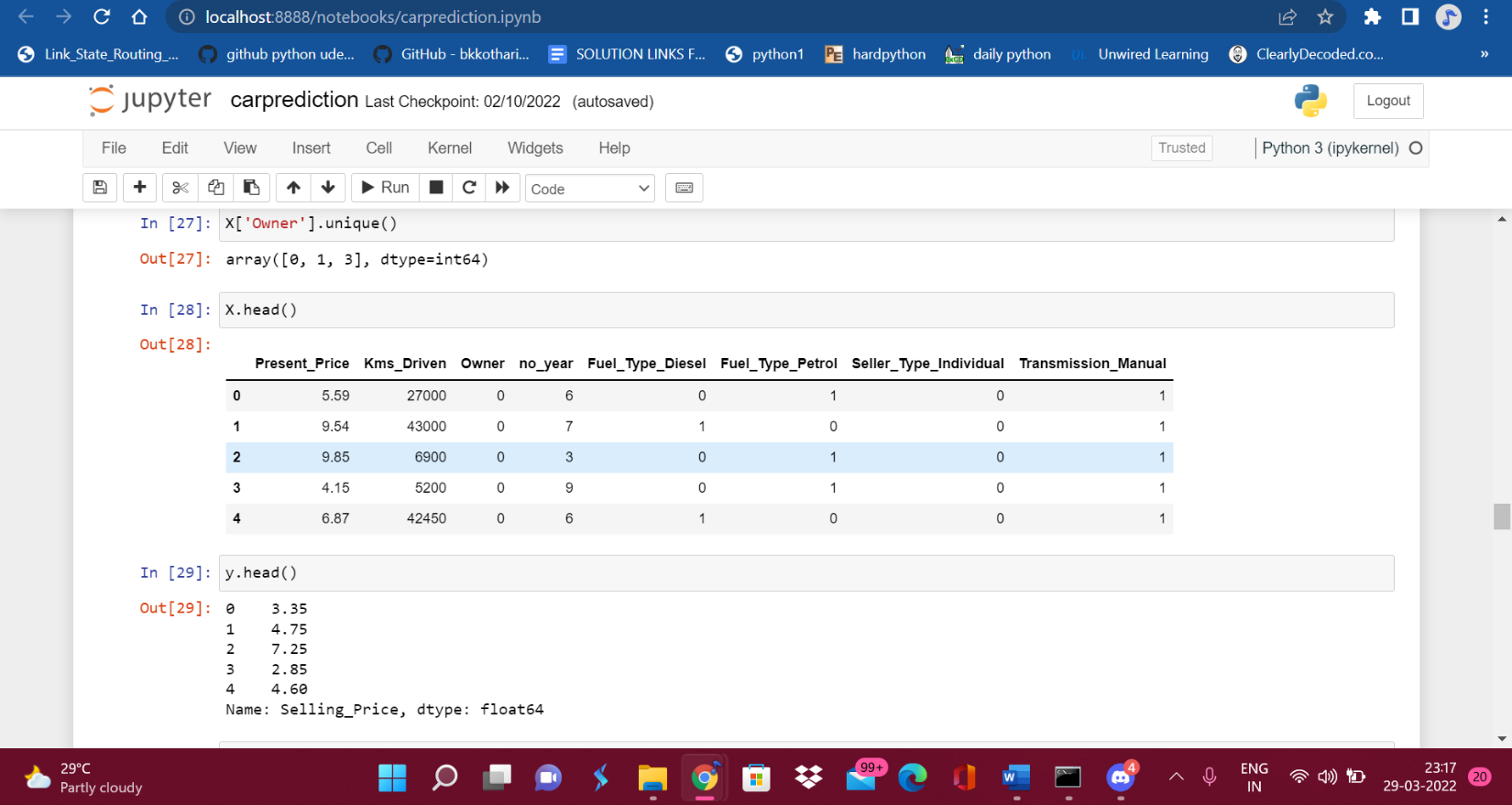


As we can see, there is very little overlap in the dataset, so we can't use knn, linear regression, or svm, and we can't even use decision trees because of the dynamic nature of the dataset, instead we'll use random forest .

**Correlations of each features in dataset:**

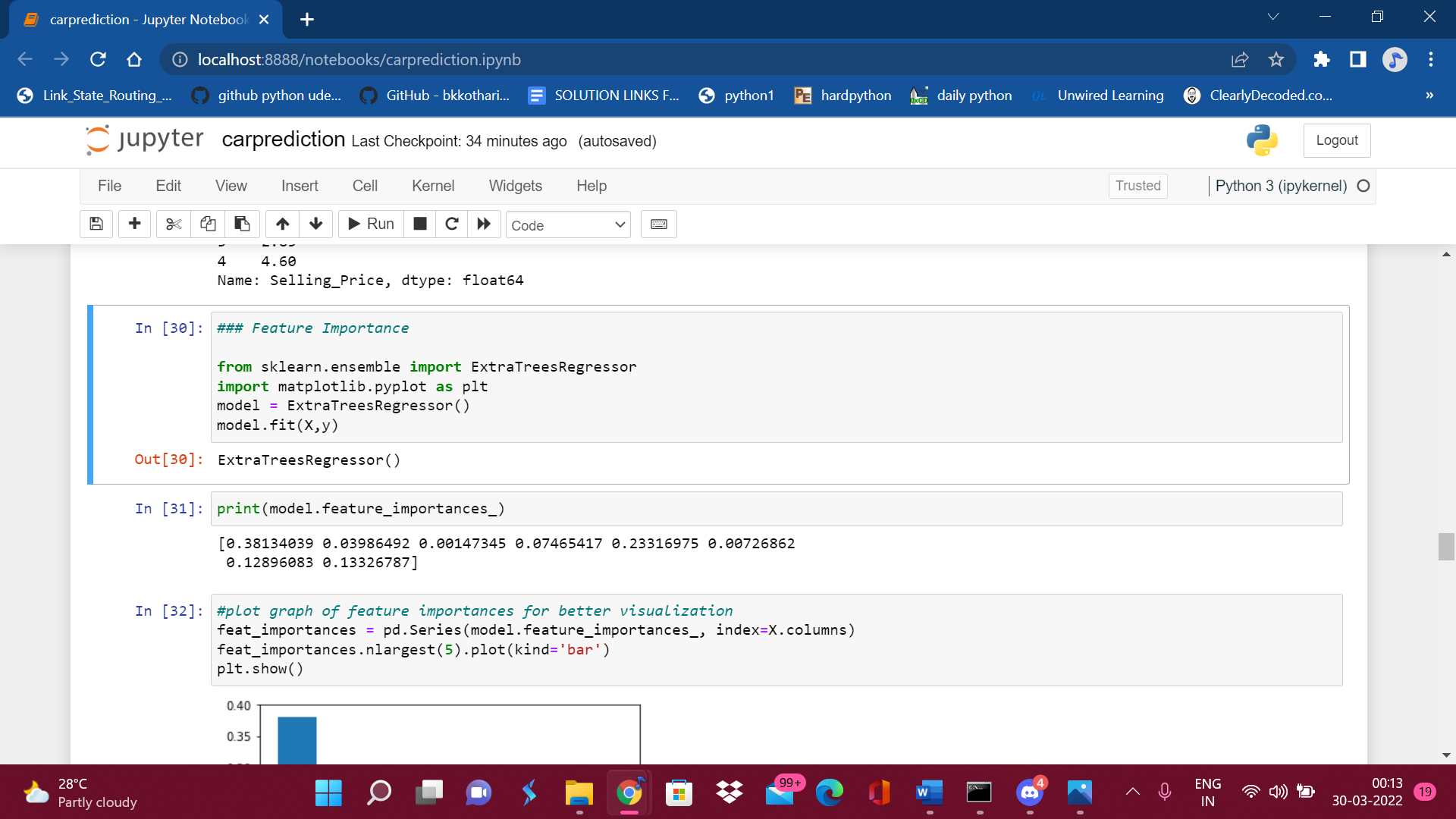


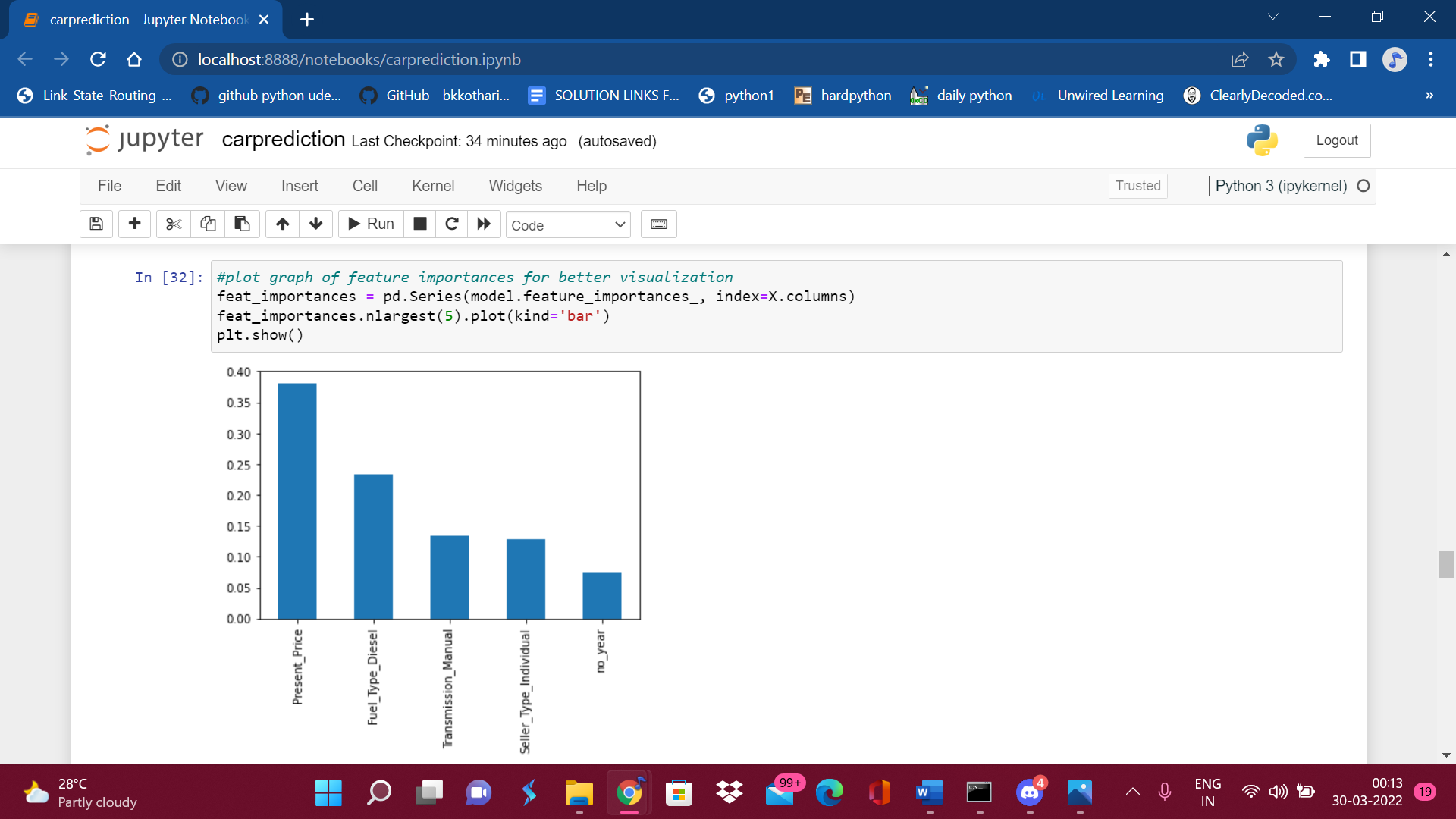


****

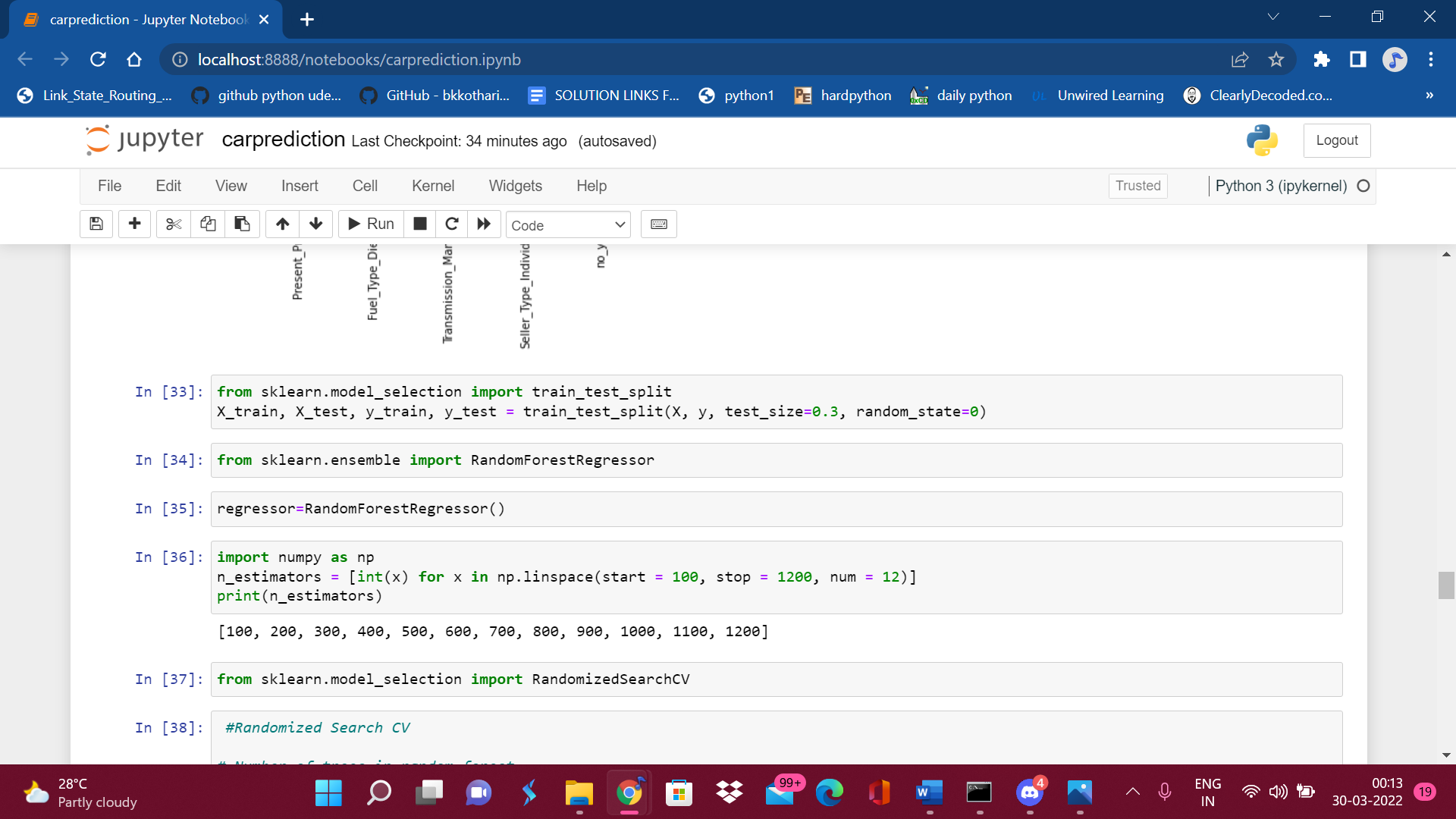
**Featured Importance:**

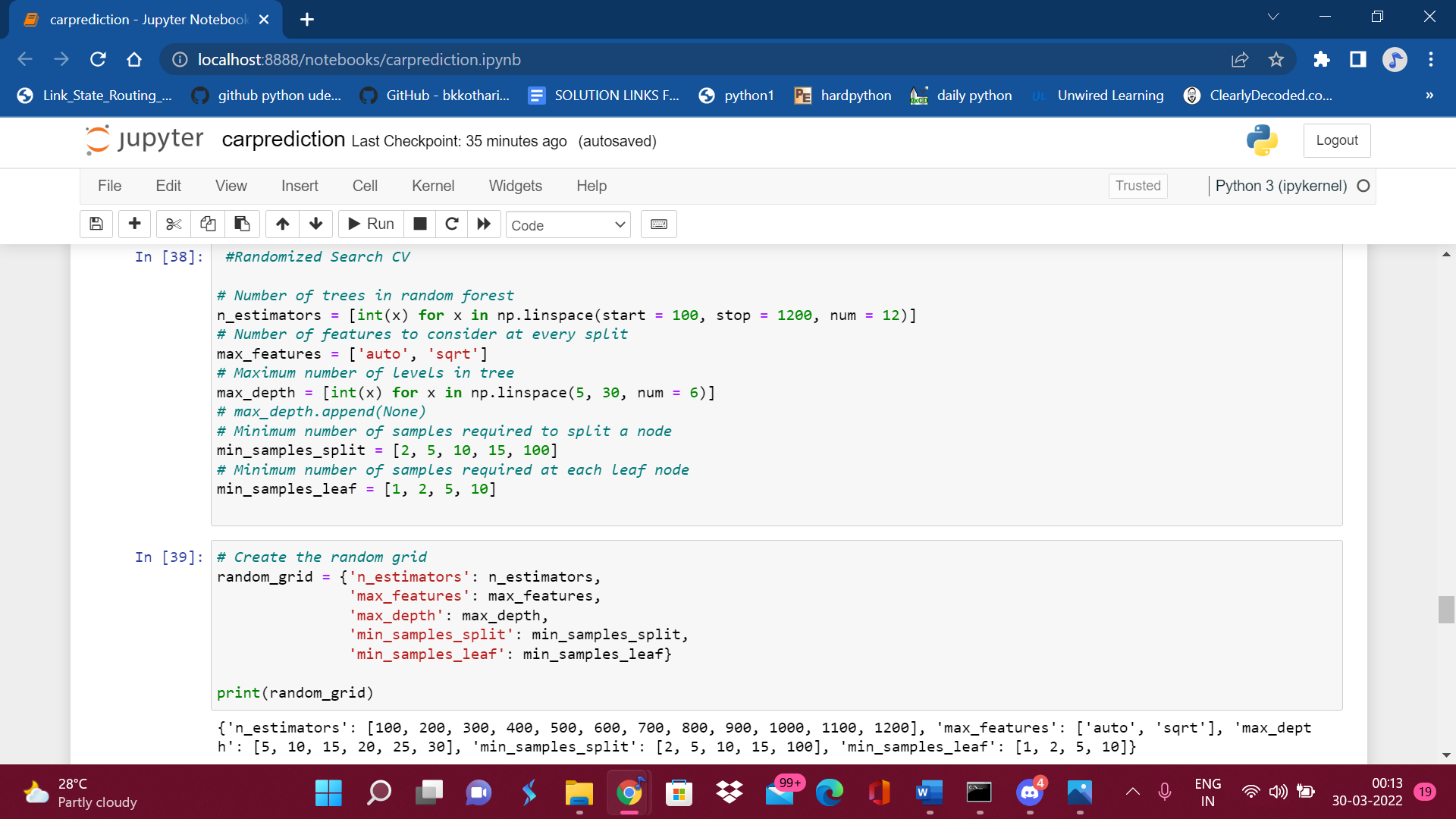
Feature Importance refers to techniques that calculate a score for all the input features for a given model- the scores simply represent the “importance” of each feature. A higher score means that the specific feature will have a larger effect on the model that is being used to predict a certain value.



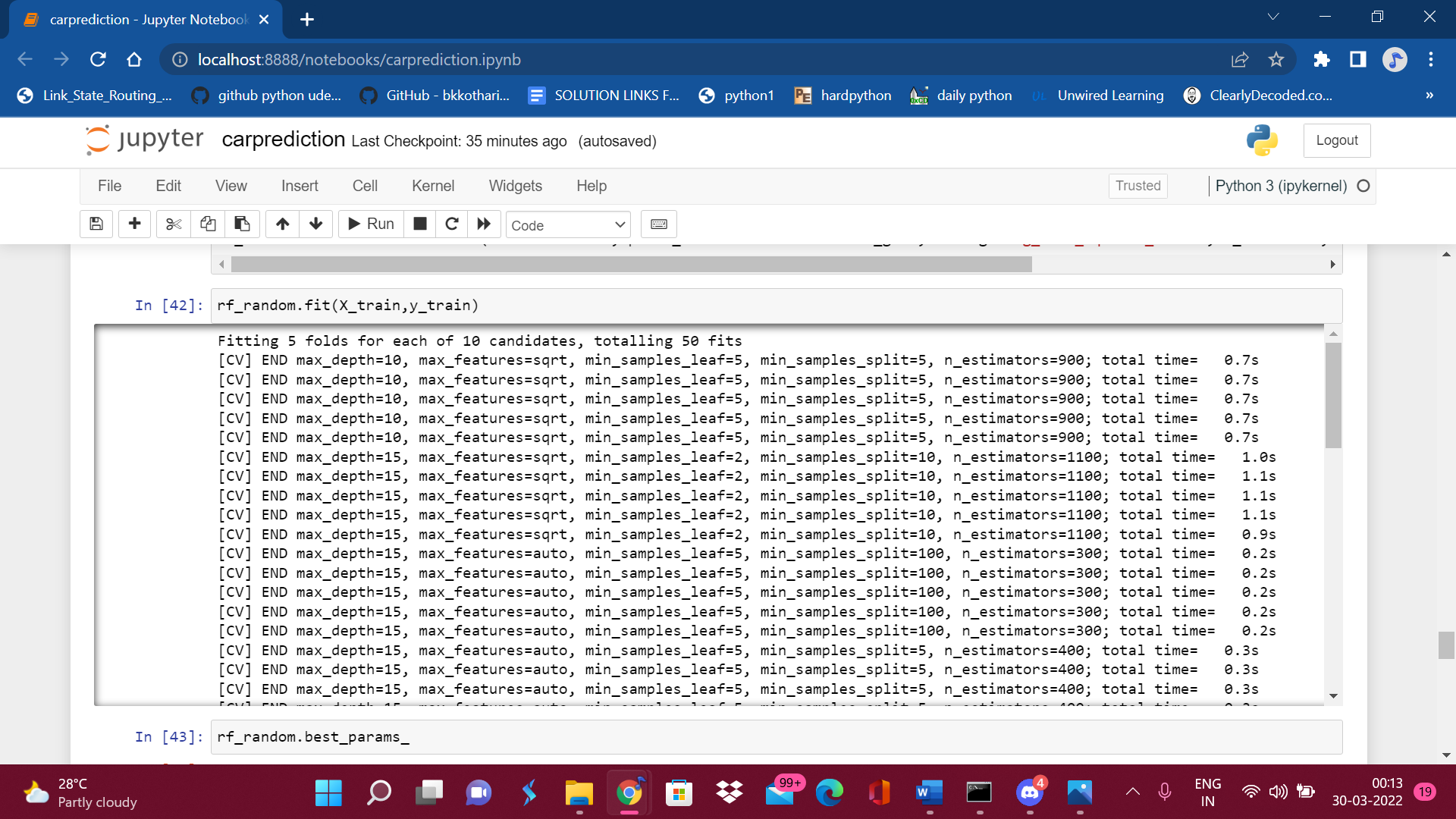


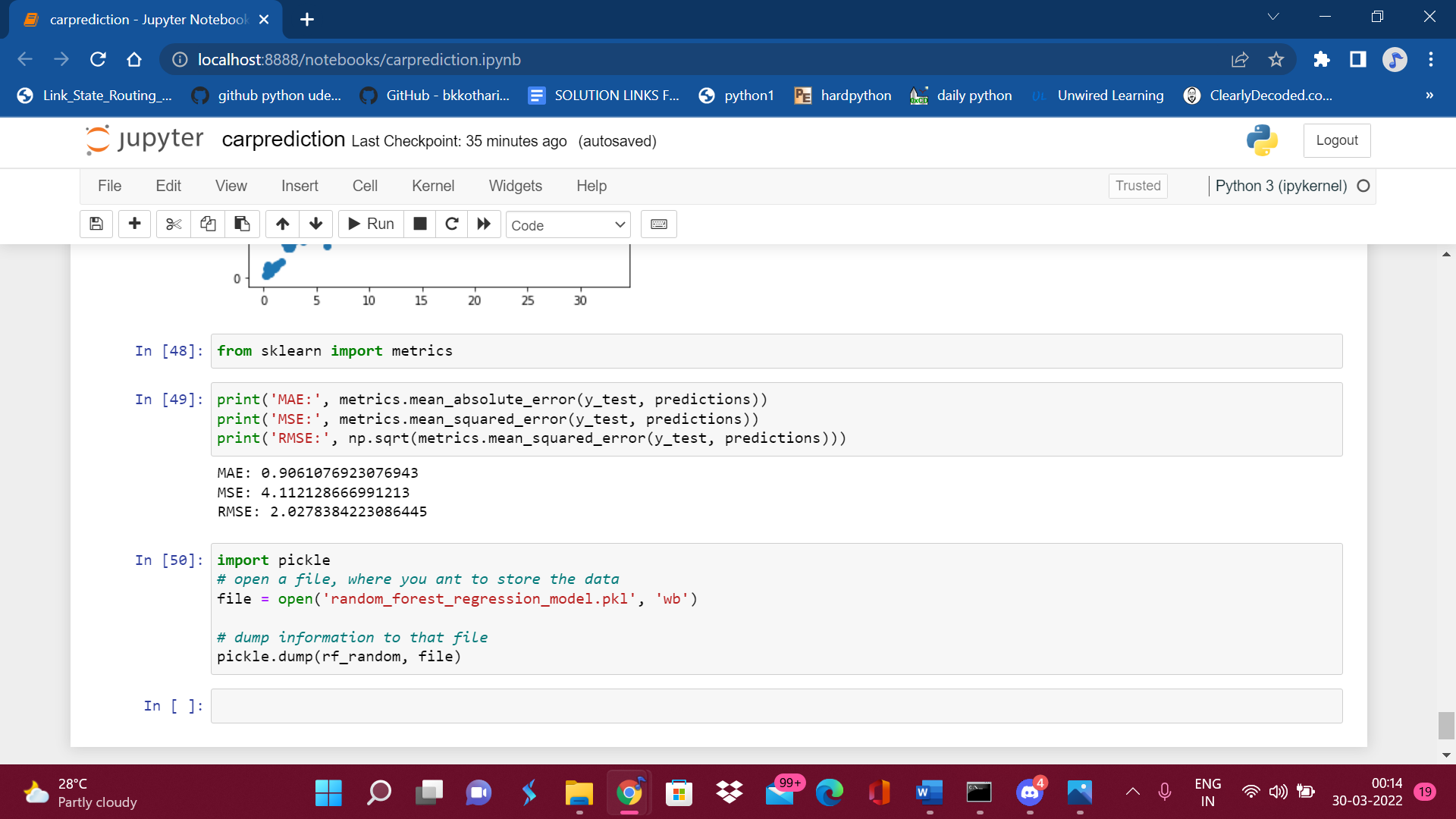
**Random forest Regressor:**





**Train and Test the Data:**

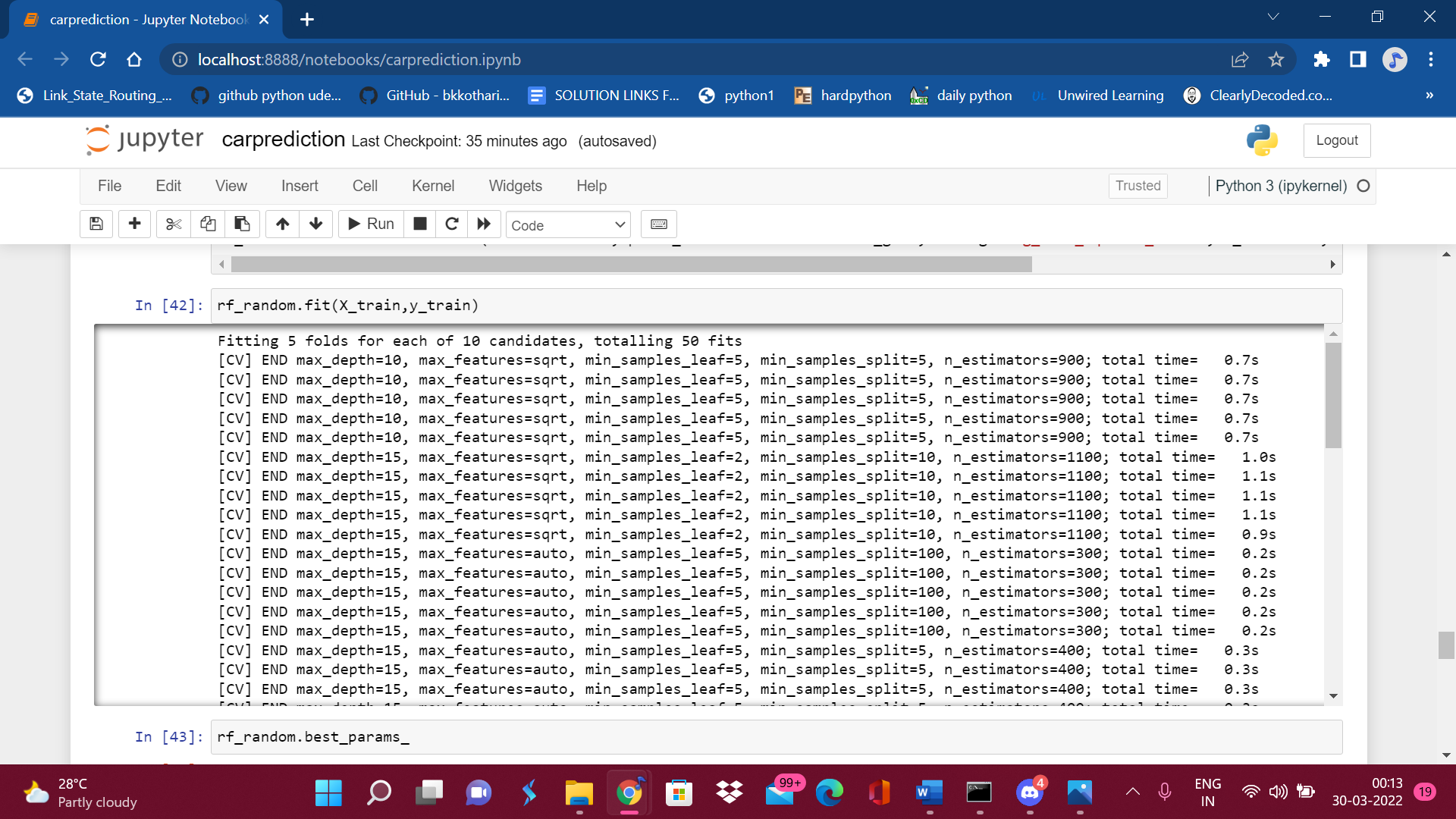


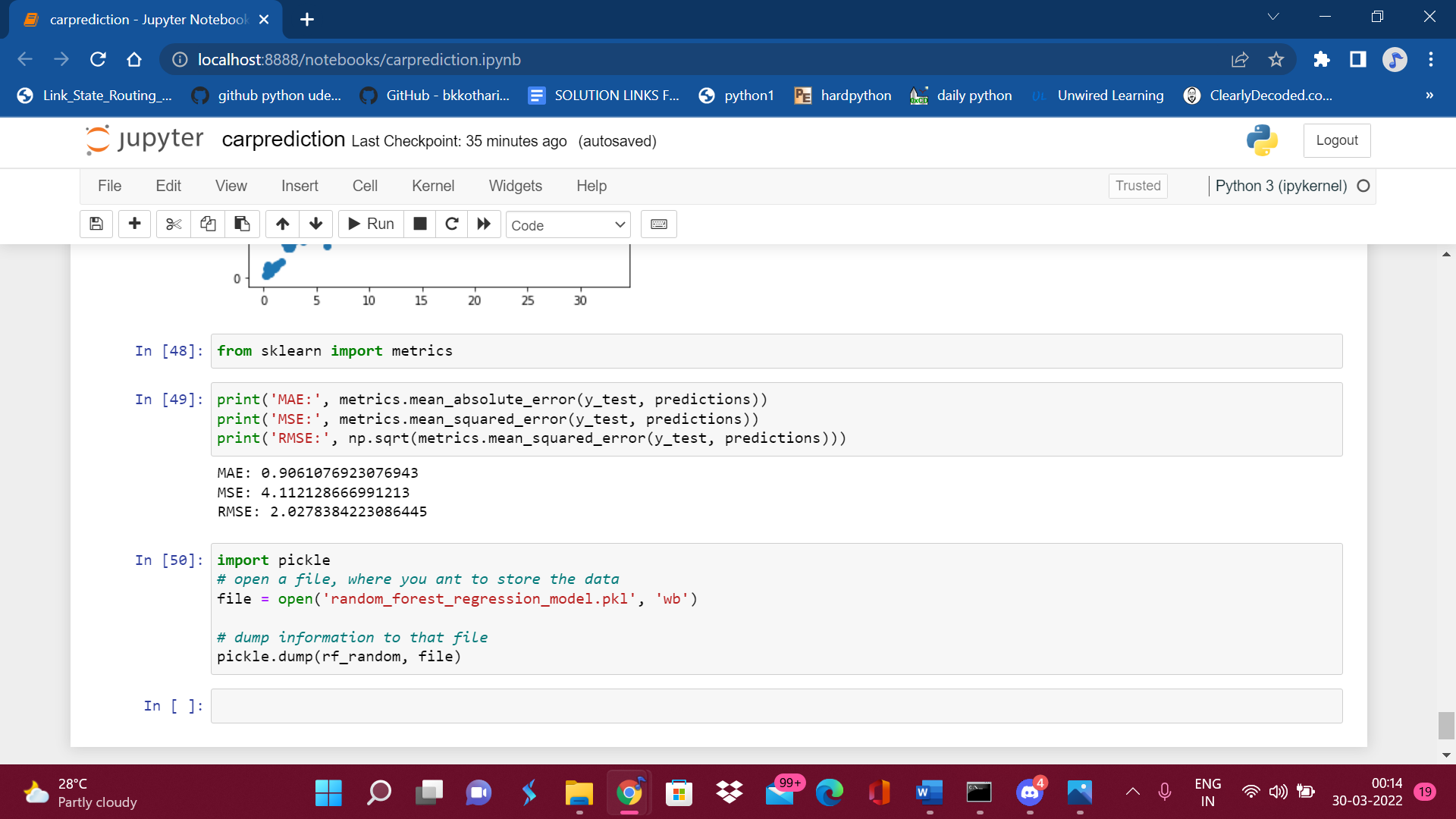


**PRE modeling steps :**

1.Splitting data set into dependent and independent variable

2.split the dataset into train and test set





**2.Testing**

checking which features are important for output features out of all the given ones.

Icon, bar chart

Description automatically generated

No\_year, Transmission\_Manual, seller\_Type\_individual, Fuel\_Type\_diesel, present\_Pricemake most impact to output prediction.

**Results and discussions:**

Chart, histogram

Description automatically generated

As we can see from the above accuracy score, The graph shows normal distribution,

We will also plot the graph using sns.displot(Y test- prediction).

Chart, scatter chart

Description automatically generated

**8.Conclusion & Future Scope**

In the future, our machine learning model can be linked to a variety of websites that give real-time data for price forecasting. We might also include a lot of previous automobile price data to assist enhance the machine learning model's accuracy. As a user interface for engaging with people, we may create an Android app. We propose to utilize intelligently designed deep learning network topologies, adjustable learning rates, and train on clusters of data rather than the entire dataset to improve performance.

The data set utilized in this study might be very useful in future studies employing other prediction approaches. This data collection may also be used to predict automobile prices using the same or other prediction tools. The data collected for this study aided in the prediction of used automobile prices using the linear regression approach. On the basis of the data set, many assumptions were made. During the processing and analysis phase, the suggested system analyzed variables and chose the most relevant variables from the dataset, reducing the complexity of the model by deleting irrelevant variables.

**9.References**

<https://medium.datadriveninvestor.com/end-to-end-project-on-used-car-price-prediction-3dc412d24aa0>

SameerchandPudaruth, “Predicting the Price of Used

Cars using Machine Learning Techniques”;

<https://www.researchgate.net/publication/319306871_Predicting_the_Price_of_Used_Cars_using_Machine_Learning_Techniques>